

TRABALHO FINAL 01 – PREÇOS DE IMÓVEIS

O trabalho visa explorar um conjunto de dados contendo informações sobre preços de imóveis com o objetivo de identificar padrões, prever preços e categorizar os imóveis de acordo com suas características.

Descrição do Conjunto de Dados:

- **Price:** O preço da casa.
- **Area:** A área total da casa em pés quadrados.
- **Bedrooms:** O número de quartos na casa.
- **Bathrooms:** O número de banheiros na casa.
- **Stories:** O número de andares na casa.
- **Mainroad:** Se a casa está conectada à estrada principal (Sim/Não).
- **Guestroom:** Se a casa tem um quarto de hóspedes (Sim/Não).
- **Basement:** Se a casa tem um porão (Sim/Não).
- **Hot water heating:** Se a casa tem um sistema de aquecimento de água quente (Sim/Não).
- **Airconditioning:** Se a casa tem um sistema de ar condicionado (Sim/Não).
- **Parking:** O número de vagas de estacionamento disponíveis na casa.
- **Prefarea:** Se a casa está localizada em uma área preferencial (Sim/Não).
- **Furnishing status:** O status de mobília da casa (Totalmente Mobiliada, Semi-Mobiliada, Sem Mobiliário).

Link para o dataset: [Housing.csv](#) ou [Kaggle](#)

FASE 1: Exploração Inicial de Dados (30 pts)

a. Limpeza e Visualização dos Dados (15 pts)

- **Identificação de Problemas:** Apresente amostras do conjunto de dados que revelem problemas como valores ausentes, inconsistências ou outliers. Documente e ilustre essas questões com exemplos concretos.
- **Técnicas Utilizadas:** Descreva e implemente as técnicas específicas de limpeza de dados que foram aplicadas, incluindo remoção de outliers, imputação de valores ausentes e correção de inconsistências. Forneça uma explicação detalhada de cada técnica utilizada.
- **Justificativa das Escolhas:** Justifique as escolhas das técnicas de limpeza aplicadas, detalhando como cada decisão contribuiu para a melhoria da qualidade dos dados e a preparação adequada para análises subsequentes.

b. Análise Exploratória (15 pts)

- **Análises Preliminares:** Realize e apresente análises detalhadas sobre as distribuições dos atributos e quaisquer padrões observados nos dados. Forneça visualizações que ajudem a ilustrar essas distribuições e padrões.
- **Correlação entre Atributos:** Investigue a existência de correlações entre os atributos do conjunto de dados e discuta as implicações dessas correlações para a modelagem futura.
- **Atributos Não Discriminativos:** Identifique e discuta os atributos que não são discriminativos ou que possuem pouca informação útil para a análise. Avalie como a presença desses atributos pode impactar a modelagem.

- **Análise de Quartis:** Execute análises de quartis para examinar as características dos imóveis em relação a atributos importantes como preço e área. Forneça insights sobre como essas características se distribuem e o que elas revelam sobre o conjunto de dados.

FASE 2: Análise Preditiva (Aprendizado de Máquina) (70 pts)

a. Identificação de Grupos de Imóveis (Clustering – K-means) (20 pts)

- **Objetivo:** Identifique pelo menos quatro grupos distintos de imóveis utilizando o algoritmo K-means. Descreva os atributos utilizados para a clusterização e analise os grupos formados.
- **Análise dos Grupos:**
 - **Características Comuns:** Identifique e explique as características comuns entre os imóveis de cada grupo.
 - **Padrão de Agrupamento:** Determine se é possível identificar um padrão de agrupamento nos dados.
 - **Impacto do Atributo "Price":** Avalie o efeito de incluir ou excluir o atributo “price” na definição dos grupos. Explique como a inclusão ou exclusão deste atributo afeta os resultados da clusterização.
 - **Ajuste de Preços:** Verifique a possibilidade de ajustar os preços dos imóveis com base nos grupos identificados, destacando aqueles que estão distantes do centro dos grupos. Apresente visualizações dos grupos e identifique imóveis que podem estar sub-precificados, sugerindo possíveis ajustes de preço.

b. Previsão de Preços (Aprendizado Supervisionado – Regressão) (20 pts)

- **Objetivo:** Avalie a capacidade de prever os preços dos imóveis utilizando diferentes modelos de regressão.
- **Comparação de Modelos:** Compare o desempenho dos diversos modelos de regressão e identifique o modelo que apresenta o melhor desempenho na previsão dos preços.

c. Classificação (Aprendizado Supervisionado – Classificação) (30 pts)

- **Categorize os Imóveis:** Classifique os imóveis em categorias como Mansões, Alto Padrão, Médio Padrão e Baixo Padrão. Utilize análises de quartis para definir essas categorias. Por exemplo, imóveis de alto padrão podem estar no primeiro e segundo quartil dos atributos área, banheiros, quartos e vagas de estacionamento, e ter características como sistema de água quente e ar-condicionado. Fique livre para definir sua estratégia de categorização.
- **Desenvolvimento e Análise dos Modelos:** Após a categorização, desenvolva três modelos de classificação distintos. Apresente e analise a performance desses modelos, incluindo matrizes de confusão e visualizações dos dados.
- **Avaliação do Impacto do "Preço":** Para o modelo com melhor desempenho, realize duas análises: uma considerando o atributo “preço” como parte do vetor de características e outra desconsiderando-o. O objetivo é avaliar se a inclusão do preço introduz viés nos modelos.

DICA: Não esqueça de realizar as normalizações dos dados. Também, se julgar necessários, utilize matrizes de correlação para verificar a representatividade dos dados. De outro modo, verifique a necessidade ou impacto de uma redução de atributos (PCA).